

A new coding theorem for three user discrete memoryless broadcast channel

Arun Padakandla and S. Sandeep Pradhan, *

Department of Electrical Engineering and Computer Science,

University of Michigan, Ann Arbor, MI 48109, USA

email: arunpr@umich.edu, pradhanv@eecs.umich.edu

August 2, 2012

Abstract

We propose a new coding technique based on nested coset codes and derive a new achievable rate region for a general three user discrete memoryless broadcast channel (DMBC). We identify an example of a three user binary broadcast channel for which the proposed achievable rate region strictly outperforms that obtained by a natural extension of Marton's [1] rate region. As a step towards deriving the achievable rate region for the general three user DMBC, we introduce the new elements of our coding theorem through a new class of broadcast channels called 3-to-1 broadcast channels.

1 Introduction

The problem of characterizing the capacity region of a broadcast channel was proposed by Cover [2] in 1972, and he introduced a novel coding technique to derive achievable rate regions for particular degraded broadcast channels. In a seminal work aimed at deriving an achievable rate region for the general degraded broadcast channel, Bergmans [3] generalized Cover's technique into what is currently referred to as superposition coding. Gallager [4] and Bergmans [5] concurrently and independently proved optimality of superposition coding for the class of degraded broadcast channels. This in particular yielded capacity region for the scalar additive Gaussian broadcast channel. However, the case of general discrete memoryless broadcast channel (DMBC) remained open. This (problem) led to the discovery of another ingenious coding technique¹. In 1979, following the works of [6, 7], Marton [1] proposed the technique of binning. In conjunction with superposition, she derived the largest known rate region for the general two user DMBC. A generalization [8, p.391 Problem 10(c)] of superposition and binning to incorporate a common message is the largest known rate region for the general DMBC and its capacity is yet unknown.²

*This work was supported by NSF grants CCF-0915619 and CCF-1116021.

¹We remark that general DMBC is richer in terms of the strategies it permits.

²It is of interest to note that though superposition and binning were known in particular settings [2], [9], its generalization led to fundamentally new ideas. For example, the description of a rate region using an auxiliary random variable [3] and the technique of binning have proved to be invaluable in deriving information theoretic achievable rate regions.

Though the capacity region has been found for many interesting classes of broadcast channels [1, 2, 10–20], the question of whether the rate region derived by Marton is optimal for the general DMBC has remained open for over thirty years. Following a period of reduced activity, there has been renewed interest [21, 22] in settling this question. Gohari and Anantharam [23] have proved computability of Marton’s rate region. This enabled them identify a class of binary broadcast channels for which Marton’s [1] rate region when computed is strictly smaller than the tightest known outer bound [24, 25], which is due to Nair and El Gamal. On the other hand, Weingarten, Steinberg and Shamai [26] have proved Marton’s binning (also referred to, in the Gaussian setting, as Costa’s dirty paper coding [27]) to be optimal for Gaussian MIMO broadcast channel, and thereby characterized capacity region for the particular class of Gaussian vector broadcast channels. It is of interest to note the optimality of Marton’s binning technique for Gaussian vector broadcast channels with arbitrary number of receivers. In this article, we (1) propose a new coding technique based on structured codes that enables us to (2) derive a new achievable rate region for the general three user discrete broadcast channel, and thereby (3) provide a strict enlargement of the current known largest achievable rate region³. In addition to providing a new coding technique and a new achievable rate region, we provide an example of a binary additive three user broadcast channel for which the proposed rate region contains a triple of rates not achievable using Marton’s technique. Indeed, one of the key elements of our work is an analytical proof of suboptimality of Marton’s rate region for the three user broadcast.

This framework proposed herein is based on our earlier work on the distributed source coding and 3-user interference channels in the discrete memoryless setting [28, 29]. The coding technique proposed herein is reminiscent of that proposed for the general three user interference channel.

While at first glance, it appears that the gains we project are similar to that harnessed in the interference channel, we opine that the phenomenon that is exploited here is fundamentally different. In a two user broadcast channel, signals intended for a user interfere with signals intended for the other. The two coding techniques - superposition and binning - exemplify the two ways interference can be tackled. Firstly, superposition enables each user decode one component of the other user’s signal and thus subtract it off. Secondly, binning enables the encoder counter each user’s interfering signal not decoded by the other by precoding for the same. Except for particular cases, the most popular being dirty paper coding, precoding results in a rate loss, i.e., in other words, precoding at the encoder is less efficient than decoding the interfering signal at the decoder. The presence of a rate loss motivates each decoder to decode as large a part as possible of the interference pattern.⁴ However decoding a large part of the interference constrains the individual rates. In a three user broadcast channel, each user’s reception is plagued by interference caused by signals intended for the other two users. The interference is in general a bi-variate function of signals intended for the other users. If the signals of the two users are endowed with structure that can help compress the range of this bi-variate function when applied to all possible signals, then the receivers can decode a large part of the interfering signal. This minimizes the component of the interference precoded, and therefore the rate loss.⁵ This is where codebooks endowed with algebraic structure outperform unstructured

³The largest known achievable rate region for the general three user discrete broadcast channel is the natural extension of Marton’s rate region for the two user case. We henceforth refer to this as Marton’s rate region for three user DMBC

⁴For the Gaussian case, there is no rate loss. Thus the encoder can precode all the interference. Indeed, the optimal strategy does not require any user to decode a part of signal not intended for it.

⁵For the Gaussian case, precoding suffers no rate loss and hence no part of the interference needs to be decoded. Thus

independent codebooks. Indeed, linear codes constrain the interference pattern to an affine subspace if the interference is the sum of user 2 and 3's signals. It is our belief that additional degrees of freedom prevalent in a three user information theoretic problem can be harnessed with codebooks endowed with algebraic structure. Whether structure in codebooks can be exploited for a two user problem remains open.

The astute reader will question the case when the bi-variate function is not a sum. Towards answering this question, we consider a natural generalization of linear codes to sets with looser algebraic structure such as groups. Our investigation of group codes, kernels of group homomorphisms, to improve achievable rate regions for information theoretic problems has been pursued in a concurrent research thread [30].⁶

The role of structured codes for improving information theoretic rate regions began with the ingenious technique of Korner and Marton [32] proposed for the source coding problem of computing modulo two sum of distributed binary sources. Han and Kobayashi [33] categorized a class of function reconstruction problems for which Korner and Marton's technique provided strict gains over the largest known rate regions using unstructured codes. Ahlswede and Han [34] proposed a universal coding technique that brings together coding techniques based on unstructured and structured codes⁷. More recently, there is a wider interest [35–37] in developing coding techniques for particular problem instances that does better than the best known techniques based on unstructured codes. It was shown in [33], in the setting of distributed source coding that for every any non-trivial and truly bi-variate function, there exists at least one source distribution for which linear codes outperform random codes. Even then, it was largely believed that codebooks possessing algebraic structure can be exploited only for modulo additive channel and source coding problems. Indeed, linear codes were known to be sub optimal for communicating over arbitrary point to point channels (and similarly for lossy compression of sources subject to an arbitrary distortion), and therefore, the basic building block in the coding scheme for any multi-terminal communication problem could not be filled by linear codes. For over thirty years, since the work of Korner and Marton came to light, neither did we know of a coding technique based on unstructured codes that did as well, nor did we know of a framework for coding based on structured codes for which the above findings was a particular case.

Krithivasan and Pradhan [28] have proposed the ensemble of nested coset codes as the basic building block of algebraic codes for compressing sources subject to any arbitrary distortion. They employ this ensemble to propose a framework for communicating information from distributed encoders observing correlated sources to a centralized decoder. Firstly, this framework generalizes the technique proposed by Korner and Marton for the general problem of distributed function computation, joint quantization of distributed sources etc. Secondly, in conjunction with the Berger Tung [38] technique this strictly enlarges

constraining interference patterns is superfluous. This explains why lattices are not necessary to achieve capacity of Gaussian vector broadcast channel.

⁶We also bring to the attention of the interested reader, our investigation [31] of pseudo group codes. While linear codes are completely *compressive* under the operation of addition, and unstructured independent codes are completely *explosive*, pseudo group codes lie in between. In other words, when two pseudo group codes of rate R are operated under the group operation, the range of the resulting codebook lies between R and $2R$. Pseudo group codes are of interest since they outperform group codes for point to point communication.

⁷Indeed, the coding techniques based on structured codes do not substitute for coding techniques based on unstructured codes. For example, reconstructing a pair of sources losslessly using two source codes that are partitioned using a common channel code can be strictly sub optimal. Similarly, the technique of partitioning independent source codes using independent channel codes is sub optimal for the problem of losslessly reconstructing modulo two sum of binary sources.

the largest known achievable rate region for the problem of distributed function computation. In the same spirit, in [29] we proposed the ensemble of nested coset codes as an ensemble of codes possessing algebraic structure that achieves capacity of arbitrary point to point channels. The technique proposed by Philosof and Zamir has been elevated to derive a new achievable rate region [39] for an arbitrary discrete multiple access channel with distributed states. We employed this ensemble to derive a new achievable rate region for the general discrete three user interference channel.

We propose a framework based on structured codes that enables us derive new achievable rate region described through information theoretic quantities for a general three user broadcast channel. Secondly, we propose the technique of joint typical encoding and decoding with codebooks possessing algebraic structure. Thirdly, our analysis of error events using correlated codebooks contains new elements.

This article is organized as follows. We begin with preliminaries in section 2. In section 3, we introduce a binary additive three user broadcast channel for which Marton's rate region is strictly sub optimal. In section 4, we define the class of 3-to-1 broadcast channels and generalize the coding technique proposed for the binary example to a general 3-to-1 broadcast channel. Finally, in section 5, we propose a coding technique for the general three user DMBC based on nested coset codes.

2 Preliminaries

A three-user discrete memoryless broadcast channel (DMBC) used without feedback is a sextuple $(\mathcal{X}, \mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3, W_{Y_1, Y_2, Y_3|X}, c)$ where \mathcal{X} is the input alphabet, \mathcal{Y}_i for $i = 1, 2, 3$ are the three output alphabets, a collection of distributions $W_{Y_1, Y_2, Y_3|X}(\cdot, \cdot, \cdot | x)$ on $\mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_3$, one for every $x \in \mathcal{X}$, and a cost function $c : \mathcal{X} \rightarrow \mathbb{R}^+$. The channel is assumed to be memoryless.

Definition 1. An $(n, \Theta_1, \Theta_2, \Theta_3)$ transmission system for a given DMBC consists of an encoder mapping

$$e : [\Theta_1] \times [\Theta_2] \times [\Theta_3] \rightarrow \mathcal{X}^n,$$

where $[\Theta_i] = \{1, \dots, \Theta_i\}$, and three decoder mappings

$$g_i : \mathcal{Y}_i^n \rightarrow \{1, \dots, \Theta_i\}$$

for $i = 1, 2$, and 3.

We assume that the messages (M_1, M_2, M_3) are drawn uniformly from the set $\{1, \dots, \Theta_1\} \times \{1, \dots, \Theta_2\} \times \{1, \dots, \Theta_3\}$. The cost associated with a vector x^n of length n is additive and is given by

$$c(x^n) = \frac{1}{n} \sum_{i=1}^n c(x_i).$$

The average error probability of the above transmission system is given by

$$\tau = \frac{1}{\Theta_1 \Theta_2 \Theta_3} \sum_{m_1=1}^{\Theta_1} \sum_{m_2=1}^{\Theta_2} \sum_{m_3=1}^{\Theta_3} \Pr((g_1(Y_1^n), g_2(Y_2^n), g_3(Y_3^n)) \neq (m_1, m_2, m_3) | (M_1, M_2, M_3) = (m_1, m_2, m_3)),$$

and the average cost incurred is given by

$$\mu = \frac{1}{\Theta_1 \Theta_2 \Theta_3} \sum_{m_1=1}^{\Theta_1} \sum_{m_2=1}^{\Theta_2} \sum_{m_3=1}^{\Theta_3} c(e(m_1, m_2, m_3)).$$

Definition 2. A quadruple of rates and cost (R_1, R_2, R_3, C) is said to be achievable for a given DMBC if $\forall \epsilon > 0$, there exists an $N(\epsilon)$ such that for all $n > N(\epsilon)$ there exists an $(n, \Theta_1, \Theta_2, \Theta_3)$ transmission system that satisfies the following conditions

$$\frac{1}{n} \log \Theta_i \geq R_i - \epsilon, \quad \text{for } i = 1, 2, 3, \quad \tau \leq \epsilon, \quad \mu \leq C + \epsilon.$$

The set of all achievable rate triples at cost C is the capacity region of the DMBC at cost level C .

3 Binary Example

We present an example of a three user DMBC for which the natural extension of the largest known rate region, due to Marton is strictly sub-optimal. In particular, we identify a triple of rates that is achievable using a coding technique based on linear codes that is not contained in Marton's rate region.

3.1 Description of the three user broadcast channel

Let $\mathbb{F}_2 = \{0, 1\}$ denote the binary field and \oplus_2 denote addition in \mathbb{F}_2 . The input alphabet is $\mathcal{X} = \mathbb{F}_2 \times \mathbb{F}_2 \times \mathbb{F}_2$ and the three output alphabets are $\mathcal{Y}_j = \mathbb{F}_2 : j = 1, 2, 3$. The channel is depicted in Fig. 1. Let $X = (X_1, X_2, X_3)$ denote the input to the channel, where $X_j \in \mathbb{F}_2$, and $Y_j \in \mathbb{F}_2$, the output at receiver j . That is, the channel has an octonary input and binary outputs. The channel transitions are described through the relations $Y_2 = X_2 \oplus_2 N_2$, $Y_3 = X_3 \oplus_2 N_3$ and $Y_1 = X_1 \oplus_2 X_2 \oplus_2 X_3 \oplus_2 N_1$, and

- N_1, N_2, N_3 are mutually independent,
- (N_1, N_2, N_3) is independent of (X_1, X_2, X_3) ,
- for $j = 2, 3$, $P(N_j = 1) = \epsilon$ and $P(N_1 = 1) = \delta$.
- $\epsilon, \delta \in (0, \frac{1}{2})$.

The input X is subject to an average cost constraint $\frac{1}{n} \mathbb{E} \{w_H(X_1^n)\} \leq q$, where w_H is the Hamming weight function, and $q \in (0, \frac{1}{2})$. We restrict to the case $q * \delta \leq \epsilon$, where $q * \delta = (1 - q)\delta + (1 - \delta)q$.

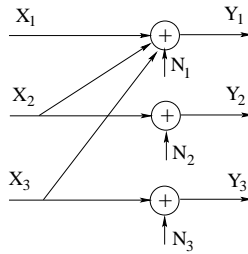


Figure 1: A 3-user broadcast channel with octonary input and binary outputs

3.2 An achievable rate region using linear codes

We present a coding technique based on linear codes that achieves the following rate region

$$\{(R_1, R_2, R_3) : R_1 \leq h_b(q * \delta) - h_b(\delta), R_2 \leq 1 - h_b(\epsilon), R_3 \leq 1 - h_b(\epsilon)\},$$

where $h_b(\cdot)$ is the binary entropy function. Let user 2 and 3 employ the same linear code that achieves capacity of a binary symmetric channel with crossover probability ϵ . User 1 employs a nested linear code [29] that achieves capacity on a binary symmetric channel with cross over probability δ and average input Hamming weight constraint q . Let $X_j^n : j = 1, 2, 3$ represent user j 's codeword. The input to the channel is $X^n = (X_1^n, X_2^n, X_3^n)$. Clearly, user 2 and 3 achieve their respective capacities. User 1 decodes $X_2^n \oplus_2 X_3^n$, the sum of user 2 and 3's transmissions. Since $q * \delta < \epsilon$, this is possible. Having decoded $X_2^n \oplus_2 X_3^n$, user 1 decodes the intended signal. It is clear that user 1 can achieve a rate $h_b(q * \delta) - h_b(\delta)$.

3.3 Sub-optimality of Marton's rate region

In this section we prove $(h_b(q * \delta) - h_b(\delta), 1 - h_b(\epsilon), 1 - h_b(\epsilon))$ is not contained in the rate region obtained by the natural extension of that proposed by Marton for the two user DMBC when $h_b(\delta * q) \leq h_b(\epsilon) < \frac{1 + h_b(\delta * q)}{2}$. In particular, we prove that if $(R_1, 1 - h_b(\epsilon), 1 - h_b(\epsilon))$ is achievable using Marton's technique, then either $R_1 < h_b(q * \delta) - h_b(\delta)$ or $R_1 = h_b(q * \delta) - h_b(\delta)$ and $h_b(\epsilon) \geq \frac{1 + h_b(\delta * q)}{2}$. Towards that end, we begin with a characterization of the rate region proposed by Marton for the two user DMBC.

Consider a two user DMBC with input alphabet \mathcal{X} , output alphabets $\mathcal{Y}_1, \mathcal{Y}_2$, channel transition probability $W_{Y_1 Y_2 | X}(\cdot, \cdot | \cdot)$ and cost function $c : \mathcal{X} \rightarrow \mathbb{R}^+$. Let $\mathbb{D}_R(W_{Y_1 Y_2 | X}, c)$ be the set of distributions⁸ $P_{W V_1 V_2 X Y_1 Y_2}$ defined over $\mathcal{W} \times \mathcal{V}_1 \times \mathcal{V}_2 \times \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2$, where $\mathcal{W}, \mathcal{V}_1$ and \mathcal{V}_2 are finite sets with $\max\{|\mathcal{W}|, |\mathcal{V}_1|, |\mathcal{V}_2|\} \leq |\mathcal{X}| + 4$ such that

- $P_{Y_1 Y_2 | X} = W_{Y_1 Y_2 | X}$,
- $W V_1 V_2 - X - Y_1 Y_2$ forms a Markov chain.

For every distribution $P_{W V_1 V_2 X Y_1 Y_2} \in \mathbb{D}_R(W_{Y_1 Y_2 | X}, c)$, let $\alpha_R(P_{W V_1 V_2 X Y_1 Y_2})$ be defined as the set of rate pairs and cost (R_1, R_2, C) such that there exists 6 non-negative real numbers K_1, K_2, S_1, T_1, S_2 and T_2 that satisfy the following constraints for $i = 1, 2$: $R_i = K_i + T_i$, and

$$\begin{aligned} 0 &\leq S_i \\ I(V_1; V_2 | W) &\leq S_1 + S_2 \\ 0 &\leq K_i \\ 0 &\leq T_i \\ I(V_i; Y_i | W) &\geq T_i + S_i \\ I(W V_i; Y_i) &\geq K_1 + K_2 + T_i + S_i, \end{aligned}$$

and $E[c(X)] \leq C$. Let $\alpha_R(W_{Y_1 Y_2 | X}, c)$ be the closure of the union of $\alpha_R(P_{W V_1 V_2 X Y_1 Y_2})$ over all distributions $P_{W V_1 V_2 X Y_1 Y_2} \in \mathbb{D}_R(W_{Y_1 Y_2 | X}, c)$.

⁸Letter R in the subscript stands for random codes

By doing Fourier-Motzkin elimination, one can easily show that $\alpha_R(P_{WV_1V_2XY_1Y_2})$ is equal to the set of all rate pairs and costs (R_1, R_2, C) such that

$$0 \leq R_1 \leq I(WV_1; Y_1) \quad (1)$$

$$0 \leq R_2 \leq I(WV_2; Y_2) \quad (2)$$

$$R_1 + R_2 \leq I(V_1; Y_1|W) + I(WV_2; Y_2) - I(V_1; V_2|W)$$

$$R_1 + R_2 \leq I(V_2; Y_2|W) + I(WV_1; Y_1) - I(V_1; V_2|W), \quad (3)$$

and $E[c(X)] \leq C$. $\alpha_R(W_{Y_1, Y_2|X}, c)$ is the Marton's achievable rate region, and is an inner bound to the capacity region.

For clarity, let us give a brief interpretation of the rate region. Let n denote the blocklength. A random code is constructed from the distribution P_W of rate $K_1 + K_2$. Let $\mathbf{W}(i)$ denote the i th codeword. For every codeword $\mathbf{W}(i)$, a code $\mathcal{C}_1(i)$ is constructed with distribution $P_{V_1|W}$ of rate $T_1 + S_1$ with $\mathbf{W}(i)$ used in the conditioning. A similar collection of codes is constructed with distribution $P_{V_2|W}$. Each V_i -code is "partitioned" into bins of rate S_i for $i = 1, 2$. Joint typical encoding and decoding is used on these codes. The standard error analysis gives the rate pairs mentioned above.

We now discuss the natural extension of Marton's rate region to a three user DMBC with input alphabet \mathcal{X} and three output alphabets $\mathcal{Y}_j : j = 1, 2, 3$ and transition probabilities $W_{\bar{Y}|X}$ ($\bar{Y} = (Y_1, Y_2, Y_3)$). Let $\alpha_R(W_{Y_1, Y_2, Y_3|X}, c)$ denote the set of all triples of rates and costs that belong to the natural extension of the Marton's rate region. The most compact description of $\alpha_R(W_{Y_1, Y_2, Y_3|X}, c)$ that we are aware of is still very long, and is given in the Appendix.

Theorem 1. *Consider the 3-receiver DMBC given in the binary example for $q, \epsilon, \delta \in (0, \frac{1}{2})$, and $q * \delta \leq \epsilon$. If $(R_1, 1 - h_b(\epsilon), 1 - h_b(\epsilon)) \in \alpha_R$, and $h_b(\epsilon) < \frac{1+h_b(q*\delta)}{2}$, then*

$$R_1 < h_b(\delta * q) - h_b(\delta).$$

Proof. See Appendix. □

It follows from the proof of this theorem, that for ϵ such that $h_b(\epsilon) \geq \frac{1+h_b(q*\delta)}{2}$, the rate point $(h_b(q * \delta) - h_b(\epsilon), 1 - h_b(\epsilon), 1 - h_b(\epsilon))$ belongs to the natural extension of the Marton's rate region.

4 3-to-1 broadcast channel

4.1 Functional Perspective on Marton's Coding

To enable us to state our new coding results succinctly with as few auxilliary random variables as possible, we will revisit the Marton's rate region for two-receiver DMBC and see how the coding can be done using coset codes. Since coset codes induce uniform single-letter distribution, we cannot perform conditional coding. In essence all codebooks are created from uniform distribution. Second, we can look at the broadcast channel from the perspective of interference channel. As noted by [40], in the two-user broadcast channel, the signals meant for the two users interfere with each other. It behooves each receiver to decode a part of the interference, i.e., a part of the signal meant for the other receiver, before decoding its own

signal. To get even better performance, they can be decoded jointly. This part of the interference can take the form of the output of a univariate function of the signal meant for the other receiver.

When we go to the three-receiver broadcast channel, we make the case for the part of the interference that is decoded at a receiver to take the form of the output of a *bivariate function* of the signals meant for the other two receivers. Consequently, group-theoretic approaches play an important role in the constructions of codes that take into account the structure of these bivariate functions. This is in contrast to the natural extension of Marton's coding approach where a pair of univariate functions of the signals meant for other two users are reconstructed at each receiver. A general approach that combines the two coding schemes can be obtained along the lines of the seminal work of Ahlswede and Han [34].

To make these concepts more concrete, consider a two-receiver DMBC $(\mathcal{X}, \mathcal{Y}_1, \mathcal{Y}_2, W_{Y_1, Y_2|X}, c)$. Let $\mathbb{D}_L(W_{Y_1, Y_2|X}, c)$ denote the set⁹ of triples $(P_{U_1, U_2, X, Y_1, Y_2}, g_1, g_2)$ of (a) probability distribution on the set $\mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2$, such that $(U_1, U_2) - X - (Y_1, Y_2)$ and $P_{Y_1, Y_2|X} = W_{Y_1, Y_2|X}$, where \mathcal{U}_1 and \mathcal{U}_2 are finite sets, and (b) two univariate functions $g_i : \mathcal{U}_i \rightarrow \mathcal{U}_i$. Let $U_{ij} = g_i(U_i)$ for $i, j = 1, 2$, and $i \neq j$. Let $\max_i |g_i(\mathcal{U}_i)| \leq p^r$, a prime power.

The first decoder jointly decodes (U_1, U_{21}) and the second decoder jointly decodes (U_2, U_{12}) . To enable the decoders decode parts of the interference, a two-level information coding procedure is employed. A code is constructed from each of the four variables U_1, U_2, U_{12} and U_{21} . This can be informally interpreted as imposing the constraint that the code of U_1 is "closed" under the univariate function $g_1(\cdot)$, and similarly for the code on U_2 .

Let us fix some notation before we proceed further. Let \mathcal{A} denote the set of all subsets of $\{1, 2, 12, 21\}$ such that (a) if 2 is present in the subset then 21 must also be present and similarly (b) if 1 is present in the subset, then 12 must also be present. When we have four real numbers, S_1, S_2, S_{12} and S_{21} (one for each element of $\{1, 2, 12, 21\}$), for all $\Theta \in \mathcal{A}$, let $S_\Theta = \sum_{a \in \Theta} S_a$ and let U_Θ is the collection $\{U_a : a \in \Theta\}$. Similarly, let \mathcal{A}_1 denote the set of all subsets of $\{1, 12, 21\}$ that contains the element 1. For $\Theta \in \mathcal{A}_1$, let Θ^c denote $\Theta^c \cap \{1, 12, 21\}$. Similarly, \mathcal{A}_2 is defined.

For every such triple, let $\alpha_L(P_{U_1, U_2, X, Y_1, Y_2}, g_1, g_2)$ be the set of all rate pairs and costs (R_1, R_2, C) such that there exists 8 non-negative real numbers S_{ij}, S_i, T_{ij} and T_i for $i, j = 1, 2$ and $i \neq j$, that satisfy the following for all $i, j = 1, 2$, and $i \neq j$: (a) $R_i = T_i - S_i + T_{ij} - S_{ij}$ (b) $T_i \geq S_i, T_{ij} \geq S_{ij}$ and $S_{ij} \leq \log p^r$, (c) covering condition for all $\Theta \in \mathcal{A}$,

$$|\Theta| \log p^r - H(U_\Theta) \leq S_\Theta$$

and (d) packing condition for all $\Theta \in \mathcal{A}_i$ for $i = 1, 2$,

$$|\Theta| \log p^r - H(U_\Theta | U_{\Theta^c}, Y_i) \geq T_\Theta$$

and $E[c(X)] \leq C$. Let $\alpha_L(W_{Y_1, Y_2|X}, c)$ denote the closure of the union of $\alpha_L(P_{U_1, U_2, X, Y_1, Y_2}, g_1, g_2)$ over all $\mathbb{D}_L(W_{Y_1, Y_2|X}, c)$.

Proposition 1. $\alpha_R(W_{Y_1, Y_2|X}, c) = \alpha_L(W_{Y_1, Y_2|X}, c)$

⁹The latter L in the subscript stands for linear codes

4.2 Coding Theorem for 3-receiver Broadcast Channel

In order to explain the proposed scheme and its novelty, we describe the same for a particular class of *3-to-1 broadcast channels*. A DMBC $(\mathcal{X}, \mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3, W_{Y_1, Y_2, Y_3|X})$ is a 3-to-1 broadcast channel if the input alphabet can be expressed as a cartesian product of three alphabets $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$ such that $W_{Y_2|X}(y_2|(x_1, x_2, x_3)) = W_{Y_2|X_2}(y_2|x_2)$ and $W_{Y_3|X}(y_3|(x_1, x_2, x_3)) = W_{Y_3|X_3}(y_3|x_3)$. Note that transition probabilities $W_{Y_1, Y_2, Y_3|X}$ of a 3-to-1 broadcast channel can be denoted as $W_{Y_1, Y_2, Y_3|X_1 X_2 X_3}$.

Following is the first main result of this paper. For a 3-to-1 broadcast channel whose transition probability is $W_{Y_1, Y_2, Y_3|X_1, X_2, X_3}$ and cost function is $c(\cdot)$, let $\mathbb{D}_L(W, c)$ denote set of triples (P, g_2, g_3) of (a) probability distribution $P_{W, U_1^3, X_1^3, Y_1^3}$ defined over $\mathcal{W} \otimes_1^3 \mathcal{U}_i \times \mathcal{X}_i \times \mathcal{Y}_i$ having the following properties

- \mathcal{U}_i is a finite set for $i = 1, 2$, and 3 , and \mathcal{W} is a finite set,
- $P_{Y_1, Y_2, Y_3|X_1, X_2, X_3} = W_{Y_1, Y_2, Y_3|X_1, X_2, X_3}$,
- $(W, U_1, U_2, U_3) - (X_1, X_2, X_3) - (Y_1, Y_2, Y_3)$ is a Markov chain

and (b) univariate functions $g_i : \mathcal{U}_i \rightarrow \mathcal{U}_i$ for $i = 2, 3$. Let p^r be a prime power such that $\max_i |g_i(\mathcal{U}_i)| \leq p^r$. Let $U_{i1} = g_i(U_i)$ for $i = 2, 3$, and let $U_{\bar{1}} = U_{21} + U_{31}$ denote the sum of U_{21} and U_{31} in the unique finite field of size p^r .

Let \mathcal{B} denote the set of all subsets of $\{1, 2, 3, 21, 31\}$ such that (a) if 2 is present in the subset then 21 must also be present and similarly (b) if 3 is present in the subset, then 31 must also be present. Let \mathcal{B}_1 denote the set of all subsets of $\{1, \bar{1}\}$ that contains the element 1, and let $T_{\bar{1}} = \max\{T_{21}, T_{31}\}$. Let \mathcal{B}_i denote the set of all subsets of $\{i, i1\}$ that contains the element i , for $i = 2, 3$.

For every triple $(P, g_1, g_2) \in \mathbb{D}_L$, let $\alpha_L(P, g_1, g_2)$ be the set of all quadruples of rates and costs (R_1, R_2, R_3, C) such that there exists 10 non-negative real numbers S_i, T_i for $i = 1, 2, 3$ and S_{i1}, T_{i1} for $i = 2, 3$ that satisfy (a) $R_1 = T_1 - S_1$, $R_2 = T_2 - S_2 + T_{21} - S_{21}$ and $R_3 = T_3 - S_3 + T_{31} - S_{31}$, (b) $T_i \geq S_i$, $\log p^r \leq S_i$ for $i = 1, 2, 3$ and $T_{i1} \geq S_{i1}$ and $\log p^r \geq S_{i1}$ for $i = 2, 3$, (c) covering condition: for all $\Theta \in \mathcal{B}$,

$$|\Theta| \log p^r - H(U_\Theta|W) \leq S_\Theta$$

and (d) packing conditions: for all $\Theta \in \mathcal{B}_i$ for $i = 1, 2, 3$,

$$|\Theta| \log p^r - H(U_\Theta|U_{\Theta^c}, Y_i, W) \geq T_\Theta$$

and $E[c(X_1, X_2, X_3)] \leq C$, and let $\alpha_L(W, c)$ denote the closure of the union of $\alpha_L(P, g_1, g_2)$ over all \mathbb{D}_L .

Theorem 2. *For every 3-to-1 DMBC (W, c) , every quadruple in $\alpha_L(W, c)$ is achievable.*

Proof. A proof will be provided in a detailed expansion of this paper. □

In the following we give an outline of the coding scheme used to achieve this rate region. We omit the formal proof in the interest of brevity.

4.3 Outline of the coding scheme

The nature of the 3-to-1 broadcast channel indicates users 2 and 3 need not decode any parts of the other users' messages. With this in mind, we propose an encoding scheme based on 5 codebooks. User 1's message is communicated using a single codebook built on \mathcal{U}_1 . User 2's message M_2 is split into two parts. The codebook built over \mathcal{U}_{21} carries message M_{21} which is a part of M_2 . The codebook built over \mathcal{U}_2 carries the rest of user 2's message. Similarly, the codebook built over \mathcal{U}_{31} carries message M_{31} which is a part of user 3's message M_3 . The codebook built over \mathcal{U}_3 carries the rest of user 3's message.

Marton's rate region involves decoding M_{21}, M_{31}, M_1 at decoder 1, M_2 at decoder 2 and M_3 at decoder 3. This involves user 1 decoding $U_{21}^n, U_{31}^n, U_1^n$, the codewords corresponding to M_{21}, M_{31} and M_1 respectively. *The key difference in the coding technique we propose is to let decoder 1 decode $U_{21}^n + U_{31}^n$ ¹⁰ instead of U_{21}^n, U_{31}^n .* While this retains uncertainty in U_{21}^n, U_{31}^n , thus enabling larger individual rates for users 2 and 3, when appropriately chosen, $U_{21}^n + U_{31}^n$ contains sufficient information of the *interference pattern*.

In order to constrain the number of possibilities for $U_{21}^n + U_{31}^n$, yet keeping the size of each codebook large, the codebooks over \mathcal{U}_{21} and \mathcal{U}_{31} are chosen to be cosets of a common linear code. Deriving source coding bounds for correlated linear codes introduces new elements. Similarly, decoding codewords from correlated channel codes introduces new elements. In the interest of brevity, we omit the details and give only the key concepts.

Fix a triple $(P, g_1, g_2) \in \mathbb{D}_L(W, c)$. We have 3 primary auxiliary random variables U_1, U_2 and U_3 , and two functions g_1 and g_2 . From these we get two secondary auxiliary random variables as $U_{21} = g_2(U_2)$ and $U_{31} = g_3(U_3)$. Generate a random sequence \mathbf{W} from the product distribution P_W^n . We will construct 5 codes, one for each random variable. The codes are built over the common finite field of size p^r . Let \mathcal{C}_i denote the code associated with the random variable U_i , and similarly let \mathcal{C}_{j1} denote the codes associated with U_{j1} for $j = 2, 3$.

The codes are constrained to be coset codes to facilitate the first decoder decode $U_{21} + U_{31}$. The codes \mathcal{C}_{21} and \mathcal{C}_{31} are nested within each other. In other words, if $|\mathcal{C}_{21}| \leq |\mathcal{C}_{31}|$, then we let $\mathcal{C}_{21} \subseteq \mathcal{C}_{31}$ and vice versa. The 5 coset codes are constructed by choosing the generator matrices randomly uniformly and independently enforcing the nesting structure between the codes of U_{21} and U_{31} . We will probabilistically partition these codes into bins. Let $\mathcal{B}_{j1}(i)$ denote the i th bin of the code associated with U_{j1} for $j = 2, 3$, and similarly, let $\mathcal{B}_j(i)$ denote the i th bin of the code associated with U_j for $j = 1, 2, 3$. The bins, however, will not have the structure of coset codes. That is, the finer codes are coset codes, and coarser codes are random codes. The reason for choosing such an ensemble is as follows. First, we only need the outer codes to be coset codes so that algebraic structure could be used to enable the first receiver decode the interference pattern. More importantly, if we let the coarse codes (bins) to be coset codes, it turns out that we lose some performance as compared to random bins.

We use letter S to denote the rates of the bins of the random variables. We use letter T to denote the rates of the codes of the random variables, respectively. For example, the rate of the bins of U_{21} is given by S_{21} , and that of U_1 is S_1 . The transmission rates are given by: $R_1 = T_1 - S_1$, $R_2 = T_2 - S_2 + T_{21} - S_{21}$, and $R_3 = T_3 - S_3 + T_{31} - S_{31}$.

Encoding: The encoder is given three messages M_1, M_2 and M_3 of rates R_1, R_2 , and R_3 , respectively,

¹⁰Recall, \mathcal{U}_{21} and \mathcal{U}_{31} are finite sets of cardinality p^r and can thus be associated with a common finite field.

to transmit. The encoder splits the second message as M_{21} and M_{22} , of rates $T_{21} - S_{21}$ and $T_2 - S_2$, respectively. Recall that $U_{21} = g_2(U_2)$ and $U_{31} = g_2(U_3)$. The encoder selects the quintuple of bins

$$(\mathcal{B}_{21}(M_{21}), \mathcal{B}_{31}(M_{31}), \mathcal{B}_1(M_1), \mathcal{B}_2(M_{22}), \mathcal{B}_3(M_{32}))$$

and looks for a quintuple of vectors, one from each that is jointly typical with \mathbf{W} with respect to the distribution $P_{W, U_{21}, U_{31}, U_1, U_2, U_3}$. If there is at least one such quintuple, then the encoder selects one of them and obtains a channel input vector by applying a random transformation of it using $P_{X|U_1, U_2, U_3}$ and sends the vector over the channel. If no such quintuple is available, the encoder declares error and sends a randomly chosen channel input vector.

It can be shown that the probability of encoding error asymptotically approaches zero if the rates of the bins are not too small. In particular, we need to have the following conditions: for all $\Theta \in \mathcal{B}$,

$$S_\Theta \geq |\Theta| \log p^r - H(U_\Theta | W)$$

Decoding: The first decoder receives the corresponding channel output vector, and looks for a unique vector pair, one from \mathcal{C}_1 and one from the larger of \mathcal{C}_{21} and \mathcal{C}_{31} , that is jointly typical with the channel output vector and the sequence \mathbf{W} with respect to the distribution $P_{W, U_1, U_{31}+U_{21}, Y_1}$. If there is such a pair, let \hat{M}_1 denote the index of the bin in \mathcal{C}_1 that contains the first vector, and declares the reconstructed message as \hat{M}_1 . Otherwise, it declares error and selects a random message for reconstruction.

The second decoder receives the corresponding channel output vector, and looks for a unique vector pair, one from \mathcal{C}_{21} and one from \mathcal{C}_2 that is jointly typical with the channel output vector and \mathbf{W} with respect to the distribution P_{W, U_{21}, U_2, Y_2} . If there exists such a pair, let \hat{M}_{21} and \hat{M}_{22} denote the indexes of the bins in \mathcal{C}_{21} and \mathcal{C}_2 that contains the unique vector pair. The decoder declares the reconstructed message as (M_{21}, M_{22}) . A similar decoding strategy is employed at the third receiver. It can be shown that the probability of decoding error at all the receivers goes to zero if the rates of the codes are not too high. In particular, we need to have for all $\Theta \in \mathcal{B}_i$ for $i = 1, 2, 3$,

$$|\Theta| \log p^r - H(U_\Theta | U_{\Theta^c} Y_i, W) \geq T_\Theta$$

5 General Broadcast Channel

5.1 Coding Theorem

In this section, we consider the general DMBC with 3 receivers. We present an achievable rate region for this channel using coset codes. This is the second main result of the paper. For a given broadcast channel described by $(W_{Y_1, Y_2, Y_3 | X}, c)$, let $\mathbb{D}_L(W, c)$ denote the set of all (a) probability distributions $P_{W, U_1, U_2, U_3, X, Y_1, Y_2, Y_3}$, defined over the sets $\mathcal{W} \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{U}_3 \times \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_3$ having the following properties:

- \mathcal{U}_i is a finite set for $i = 1, 2, 3$ and \mathcal{W} is a finite set,
- $(W, U_1, U_2, U_3) - X - (Y_1, Y_2, Y_3)$ is a Markov chain

and (b) six functions $g_{ij} : \mathcal{U}_i \rightarrow \mathcal{U}_i$ for all $i, j \in \{1, 2, 3\}$, $i \neq j$. Let $U_{ji} = g_{ji}(U_i)$. Let $p_i^{r_i}$ be a prime power such that $|\mathcal{U}_i| \leq p_i^{r_i}$, and

$$\max_{\{j \in \{1, 2, 3\} : j \neq i\}} |g_{ji}(\mathcal{U}_j)| \leq p_i^{r_i},$$

for $i = 1, 2, 3$. Hence can we endow the alphabets of U_{21}, U_{31} , and U_1 with the algebraic structure of the unique finite field of size $p_1^{r_1}$, and similarly for the triples (U_{12}, U_{32}, U_2) , and (U_{13}, U_{23}, U_3) .

We will use the following notation for a compact description of the rate region. We will use double indexed rates such as S_{ij} and double indexed random variables such as U_{ij} , where each index can take values in $I \triangleq \{1, 2, 3\}$ and $i \neq j$. For a given pair (i, j) let k denote the element in I such that $k \neq i$ and $k \neq j$. For example when $(i, j) = (1, 3)$, $k = 2$. Let $U_{i\bar{i}}$ denote the collection of random variables $\{U_{ij}, U_{ik}\}$. Let $S_{i\bar{i}}$ denote the sum rate $S_{ij} + S_{ik}$. For example $S_{1\bar{1}} = S_{13} + S_{12}$. Let $U_{\bar{i}}$ denote the sum $U_{ji} + U_{ki}$, where $+$ is the addition operation of the corresponding finite field. Let $T_{\bar{i}}$ denote $\max\{T_{ji}, T_{ki}\}$. For example $T_{\bar{1}} = \max\{T_{21}, T_{31}\}$. Observe that in $\bar{\cdot}$ notation, the index i becomes the second index. For every element $s \in \{1, 2, 3, 12, 13, 21, 23, 31, 32, \bar{1}, \bar{2}, \bar{3}\}$, let A_s denote the size of the finite field associated with the alphabet of U_s . A similar notation is used for subsets. Let π_i be a permutation on the set $\{ij, ik, \bar{i}\}$, for $i = 1, 2, 3$. Let \mathcal{B} denote the set of all subsets of $\{1, 2, 3, 12, 13, 21, 23, 31, 32\}$ such that if i is present in the subset, then ij and ik must also be present. A curious reader may note that $|\mathcal{B}| = 214$.

For every PMF P and six functions $g_{ij}(\cdot)$ in \mathbb{D}_L , let $\alpha_L(P, g_{ij})$ denote the set of all quadruples of rates and costs (R_1, R_2, R_3, C) such that there exists 18 non-negative real numbers T_i, S_i, T_{ij}, S_{ij} for all $i, j \in \{1, 2, 3\}$, $i \neq j$, and 3 permutations π_1, π_2 and π_3 , that satisfy the following for all $i, j \in \{1, 2, 3\}$, $i \neq j$, (a) $R_i = T_i - S_i + T_{i\bar{i}} - S_{i\bar{i}}$, (b) $T_{ij} \geq S_{ij}$, $T_i \geq S_i$, $S_{ij} \leq A_{ij}$ and $S_i \leq A_i$, (c) covering conditions: for all $\Theta \in \mathcal{B}$,

$$A_\Theta - H(U_\Theta|W) \leq S_\Theta$$

and (d) packing conditions

$$\begin{aligned} T_i &\leq A_i - H(U_i|U_{i\bar{i}}, U_{\bar{i}}, Y_i, W) \\ T_i + T_{\pi_i(ij)} &\leq A_i + A_{\pi_i(ij)} - H(U_i, U_{\pi_i(ij)}|U_{\pi_i(ik)}, U_{\pi_i(\bar{i})}, Y_i, W) \\ T_{\pi_i(ik)} &\leq A_{\pi_i(ik)} - H(U_{\pi_i(ik)}|U_{\pi_i(\bar{i})}, Y_i, W) \\ T_{\pi_i(\bar{i})} &\leq A_{\pi_i(\bar{i})} - H(U_{\pi_i(\bar{i})}, Y_i, W) \end{aligned}$$

and $E[c(X)] \leq C$. Let $\alpha_L(W, c)$ denote the convex closure of the union of $\alpha_L(P, g_{ij})$ over all \mathbb{D}_L .

Theorem 3. *For every DMBC (W, c) , every quadruple (R_1, R_2, R_3, C) in α_L is achievable.*

Proof. A proof will be provided in a detailed expansion of this paper. □

5.2 Coding Scheme

Each receiver is assigned a main auxilliary random variable. For example U_i for receiver i . The first receiver wishes to reconstruct a bivariate function of a part of the signal meant for receiver 2 and a part of the signal meant for receiver 3. This is given by $U_{\bar{1}} = U_{21} + U_{31} = g_{21}(U_2) + g_{31}(U_3)$. Similarly, $U_{\bar{2}} = U_{12} + U_{32} = g_{12}(U_2) + g_{32}(U_3)$ is decoded at receiver 2, and so on. We employ a form of successive

decoding strategy at each receiver. The first decoder attempts to recover the following $(U_{12}, U_{13}, U_{\tilde{1}}, U_1)$. A permutation of $\{12, 13, \tilde{1}\}$ is chosen. For example $(13, \tilde{1}, 12)$. Then U_{13} is decoded first, and then $U_{\tilde{1}}$ is decoded next, and then the pair (U_{12}, U_1) is decoded jointly. Each random variable is assigned a nested code. The fine code is a coset code, and a coarse code is obtained by “probabilistic” partitioning of the fine code. The codes associated with U_{12} , U_{32} and U_2 are constructed on the unique finite field of size $p_2^{r_2}$, and those associated with U_{13} , U_{23} and U_3 are constructed on the unique finite field of size $p_3^{r_3}$, and so on. The fine codes of U_{12} and U_{32} are nested within each other depending on their sizes. The codes associated with U_i are generated independently of the other codes.

Acknowledgements

We thank Aria Sahebi and Farhad Shirani for several discussions on this topic. The first author thanks Lalitha for providing computing resources that aided preparation of a preliminary draft of this article.

Appendix

Proof. (Theorem 1):

Let $\mathbb{D}_R(W_{Y_1 Y_2 | X}, c)$ be the set of distributions $P_{W\bar{U}\bar{V}X\bar{Y}}$ defined on $\mathcal{W} \times \bar{\mathcal{U}} \times \bar{\mathcal{V}} \times \mathcal{X} \times \bar{\mathcal{Y}}$, where

1. $\bar{\mathcal{Y}} = \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_3$
2. \mathcal{W} is a finite set
3. $\bar{\mathcal{U}} = \mathcal{U}_{12} \times \mathcal{U}_{31} \times \mathcal{U}_{23}$ and $\bar{\mathcal{V}} = \mathcal{V}_1 \times \mathcal{V}_2 \times \mathcal{V}_3$ are cartesian product of three finite sets¹¹

such that,

1. $W\bar{U}\bar{V} - X - \bar{Y}$ is a Markov chain,
2. $P_{\bar{Y}|X} = W_{\bar{Y}|X}$.

Let $\alpha_R(P_{W\bar{U}\bar{V}X\bar{Y}})$ be the set of rate triples and costs¹² (R_1, R_2, R_3, C) such that there exists 18 non-negative real numbers $K_1, K_2, K_3, K_{12}, L_{12}, S_{12}, K_{31}, L_{31}, S_{31}, K_{23}, L_{23}, S_{23}$ and $T_1, S_1, T_2, S_2, T_3, S_3$ that satisfy the following 3 rate splitting constraints $R_1 = K_1 + K_{12} + L_{31} + T_1$, $R_2 = K_2 + K_{23} + L_{12} + T_2$,

¹¹Bounds on the cardinalities of $\bar{\mathcal{U}}$ and $\bar{\mathcal{V}}$ and \mathcal{W} are not available in the literature yet.

¹²This is the three user analogue of $\alpha_R(P_{WV_1V_2XY_1Y_2})$.

and $R_3 = K_3 + K_{31} + L_{23} + T_3$, 11 covering constraints

$$\begin{aligned}
I(U_{12}; U_{23}|W) &\leq S_{12} + S_{23} \\
I(U_{12}; U_{31}|W) &\leq S_{12} + S_{31} \\
I(U_{23}; U_{31}|W) &\leq S_{23} + S_{31} \\
I(U_{12}; U_{23}; U_{31}|W) &\leq S_{12} + S_{23} + S_{31} \\
I(U_{12}; U_{23}; U_{31}|W) + I(V_1; U_{23}|U_{12}U_{31}W) &\leq S_1 + S_{12} + S_{23} + S_{31} \\
I(U_{12}; U_{23}; U_{31}|W) + I(V_2; U_{31}|U_{23}U_{12}W) &\leq S_2 + S_{12} + S_{23} + S_{31} \\
I(U_{12}; U_{23}; U_{31}|W) + I(V_3; U_{12}|U_{31}U_{23}W) &\leq S_3 + S_{12} + S_{23} + S_{31} \\
I(V_1; U_{23}|U_{12}U_{31}W) + I(V_2; U_{31}|U_{12}U_{23}W) + I(V_1; V_2|U_{12}U_{23}U_{31}W) &\leq S_1 + S_2 + S_{12} + S_{23} + S_{31} \\
&\quad - I(U_{12}; U_{23}; U_{31}|W) \\
I(V_2; U_{31}|U_{23}U_{12}W) + I(V_3; U_{12}|U_{23}U_{31}W) + I(V_2; V_3|U_{12}U_{23}U_{31}W) &\leq S_2 + S_3 + S_{12} + S_{23} + S_{31} \\
&\quad - I(U_{12}; U_{23}; U_{31}|W) \\
I(V_3; U_{12}|U_{31}U_{23}W) + I(V_1; U_{23}|U_{31}U_{12}W) + I(V_3; V_1|U_{12}U_{23}U_{31}W) &\leq S_3 + S_1 + S_{12} + S_{23} + S_{31} \\
&\quad - I(U_{12}; U_{23}; U_{31}|W) \\
I(V_1; U_{23}|U_{12}U_{31}W) + I(V_2; U_{31}|U_{12}U_{23}W) + I(V_3; U_{12}|U_{31}U_{23}W) &\leq S_1 + S_2 + S_3 + S_{12} + S_{23} + S_{31} \\
&\quad - I(V_1; V_2; V_3|U_{12}U_{23}U_{31}W) \\
&\quad - I(U_{12}; U_{23}; U_{31}|W)
\end{aligned}$$

and 15 packing constraints.

$$\begin{aligned}
I(V_1; Y_1|WU_{12}U_{31}) &\geq T_1 + S_1 \\
I(U_{12}V_1; Y_1|WU_{31}) + I(U_{12}; U_{31}|W) &\geq K_{12} + L_{12} + S_{12} + T_1 + S_1 \\
I(U_{31}V_1; Y_1|WU_{12}) + I(U_{12}; U_{31}|W) &\geq K_{31} + L_{31} + S_{31} + T_1 + S_1 \\
I(U_{12}U_{31}V_1; Y_1|W) + I(U_{12}; U_{31}|W) &\geq K_{12} + L_{12} + S_{12} + K_{31} + L_{31} + S_{31} + T_1 + S_1 \\
I(WU_{12}U_{31}V_1; Y_1) + I(U_{12}; U_{31}|W) &\geq K_1 + K_2 + K_3 + K_{12} + L_{12} + S_{12} + K_{31} + L_{31} + S_{31} + T_1 + S_1 \\
\\
I(V_2; Y_2|WU_{23}U_{12}) &\geq T_2 + S_2 \\
I(U_{12}V_2; Y_2|WU_{23}) + I(U_{12}; U_{23}|W) &\geq K_{12} + L_{12} + S_{12} + T_2 + S_2 \\
I(U_{23}V_2; Y_2|WU_{12}) + I(U_{12}; U_{23}|W) &\geq K_{23} + L_{23} + S_{23} + T_2 + S_2 \\
I(U_{12}U_{23}V_2; Y_2|W) + I(U_{12}; U_{23}|W) &\geq K_{12} + L_{12} + S_{12} + K_{23} + L_{23} + S_{23} + T_2 + S_2 \\
I(WU_{12}U_{23}V_2; Y_2) + I(U_{12}; U_{23}|W) &\geq K_1 + K_2 + K_3 + K_{12} + L_{12} + S_{12} + K_{23} + L_{23} + S_{23} + T_2 + S_2 \\
\\
I(V_3; Y_3|WU_{23}U_{31}) &\geq T_3 + S_3 \\
I(U_{31}V_3; Y_3|WU_{23}) + I(U_{23}; U_{31}|W) &\geq K_{31} + L_{31} + S_{31} + T_3 + S_3 \\
I(U_{23}V_3; Y_3|WU_{31}) + I(U_{23}; U_{31}|W) &\geq K_{23} + L_{23} + S_{23} + T_3 + S_3 \\
I(U_{23}U_{31}V_3; Y_3|W) + I(U_{23}; U_{31}|W) &\geq K_{23} + L_{23} + S_{23} + K_{31} + L_{31} + S_{31} + T_3 + S_3 \\
I(WU_{23}U_{31}V_3; Y_3) + I(U_{23}; U_{31}|W) &\geq K_1 + K_2 + K_3 + K_{23} + L_{23} + S_{23} + K_{31} + L_{31} + S_{31} + T_3 + S_3
\end{aligned}$$

Let $\alpha_R(W_{\bar{Y}|X}, c)$ be the *convex* closure¹³ of the union $\alpha_R(P_{W\bar{U}\bar{V}X\bar{Y}})$ over all distributions $P_{W\bar{U}\bar{V}X\bar{Y}} \in \mathbb{D}_R(W_{\bar{Y}|X}, c)$. The characterization of $\alpha_R(P_{W\bar{U}\bar{V}X\bar{Y}})$ can be obtained by identifying the covering and packing bounds in terms of rates of the 7 codebooks, eliminating the variables that are not of interest using the technique of Fourier-Motzkin and expressing the rate region in terms of the three rate parameters R_1, R_2, R_3 . However this procedure turns out to be cumbersome for the three user channel and yields in excess of 100 inequalities that are needed to characterize $\alpha_R(P_{W\bar{U}\bar{V}X\bar{Y}})$. The lack of a compact characterization of $\alpha_R(P_{W\bar{U}\bar{V}X\bar{Y}})$ is one of the key difficulties in establishing the sub-optimality of Marton's rate region. We circumvent this difficulty as follows. Suppose $(R_1, 1 - h_b(\epsilon), 1 - h_b(\epsilon)) \in \alpha_R(W_{\bar{Y}|X}, c)$, then there must exist a distribution, say $P_{W\bar{U}\bar{V}X\bar{Y}} \in \mathbb{D}_R(W_{\bar{Y}|X}, c)$ such that $(R_1, 1 - h_b(\epsilon), 1 - h_b(\epsilon)) \in \alpha_R(P_{W\bar{U}\bar{V}X\bar{Y}})$. This is because of the extremal nature of the operating point in the second and the third coordinates. That is, $R_2 = 1 - h_b(\epsilon)$ is the the maximum rate at which communication can take place between the encoder and the second decoder, and similarly for the third decoder. Hence $(R_1, 1 - h_b(\epsilon), 1 - h_b(\epsilon))$ cannot be a convex combination of points where the second or the third coordinate is strictly larger than or strictly smaller than $1 - h_b(\epsilon)$.

Lemma 1. *If $(R_1, 1 - h_b(\epsilon), 1 - h_b(\epsilon)) \in \alpha_R(P_{W\bar{U}\bar{V}X\bar{Y}})$, then the following conditions must be satisfied.*

$$K_1 = K_2 = K_3 = K_{23} = L_{23} = K_{12} = L_{31} = S_2 = S_3 = S_{23} = 0 \quad (4)$$

$$(Y_2, X_2, V_2, U_{12}) - (WU_{23}) - (V_3, U_{31}, X_3, Y_3) \quad (5)$$

$$(V_3, X_3, V_1, V_{13}) - (WU_{23}U_{12}V_2) - (X_2, Y_2) \quad (6)$$

$$(V_2, X_2, V_1, V_{12}) - (WU_{23}U_{31}V_3) - (X_3, Y_3) \quad (7)$$

$$S_{12} = I(U_{12}; U_{23}|W), \text{ and } S_{31} = I(U_{31}; U_{23}|W) \quad (8)$$

and (WU_{23}) is independent of Y_2 and is independent of Y_3 .

Using these relations, we can now simplify the 3 rate splitting, 11 covering and 15 packing constraints as follows: $R_2 = R_3 = 1 - h_b(\epsilon) = I(V_2U_{12}; Y_2|WU_{23}) = I(V_3U_{31}; Y_3|WU_{23})$, $R_2 \geq T_2$, $R_3 \geq T_3$,

$$I(V_1; U_{23}V_1V_2|WU_{12}U_{31}) \leq S_1$$

$$I(V_2; Y_2|WU_{12}U_{23}) \geq T_2 \geq 0$$

$$I(V_3; Y_3|WU_{31}U_{23}) \geq T_3 \geq 0$$

$$I(V_1; Y_1U_{23}|WU_{12}U_{31}) \geq R_1 + S_1$$

$$I(V_1U_{12}; Y_1U_{23}|WU_{31}) + I(U_{12}; U_{31}|W) - I(U_{23}; U_{12}|W) \geq R_2 - T_2 + R_1 + S_1$$

$$I(V_1U_{31}; Y_1U_{23}|WU_{12}) + I(U_{12}; U_{31}|W) - I(U_{23}; U_{31}|W) \geq R_3 - T_3 + R_1 + S_1$$

$$I(V_1U_{12}U_{31}; Y_1U_{23}|W) + I(U_{12}; U_{31}|W) - I(U_{23}; U_{12}|W) - I(U_{23}; U_{31}|W) \geq R_2 - T_2 + R_3 - T_3 + R_1 + S_1,$$

where we have added the following 4 non-negative quantities to the left hand sides of the last 4 equations, respectively: $I(V_1; U_{23}|Y_1WU_{31})$, $I(V_1U_{12}; U_{23}|Y_1WU_{31})$, $I(V_1U_{31}; U_{23}|Y_1WU_{12})$, and $I(V_1U_{12}U_{13}; U_{23}|Y_1W)$ which will only weaken the constraint and in turn enlarges the Marton's rate region.

¹³It is not yet clear whether the closure of the union is convex or not.

Now we are ready for a Fourier-Motzkin elimination. After such an elimination, we finally get the following rate region. $R_2 = R_3 = 1 - h_b(\epsilon) = I(V_2 U_{12}; Y_2 | W U_{23}) = I(V_3 U_{31}; Y_3 | W U_{23})$, and

$$R_1 \leq I(V_1; Y_1 | W U_{12} U_{23} U_{31}) - I(V_1; V_2 V_3 | W U_{12} U_{23} U_{31}) \quad (9)$$

$$R_1 \leq I(V_1; Y_1 | W U_{12} U_{23} U_{31}) - I(V_1; V_2 V_3 | W U_{12} U_{23} U_{31}) + I(U_{12}; Y_1 | W U_{23} U_{31}) - I(U_{12}; Y_2 | W U_{23}) \quad (10)$$

$$R_1 \leq I(V_1; Y_1 | W U_{12} U_{23} U_{31}) - I(V_1; V_2 V_3 | W U_{12} U_{23} U_{31}) + I(U_{31}; Y_1 | W U_{23} U_{12}) - I(U_{31}; Y_3 | W U_{23}) \quad (11)$$

$$R_1 \leq I(V_1; Y_1 | W U_{12} U_{23} U_{31}) - I(V_1; V_2 V_3 | W U_{12} U_{23} U_{31}) + I(U_{12} U_{31}; Y_1 | W U_{23}) - I(U_{12}; Y_2 | W U_{23}) - I(U_{31}; Y_3 | W U_{23}) \quad (12)$$

Now let us look at the first equation (equation 9) in the above four. Using the Markov chains $(V_3, X_3, V_1, U_{13}) - (W U_{23} U_{12} V_2) - (X_2, Y_2)$, and $(V_2, X_2, V_1, U_{12}) - (W U_{23} U_{31} V_3) - (X_3, Y_3)$, and denoting quadruple $(W U_{12} U_{23} U_{31})$ as \tilde{W} , and the sum $X_2 + X_3$ as S_1 , we get

$$R_1 \leq I(V_1; Y_1 | W U_{12} U_{23} U_{31}) - I(V_1; V_2 V_3 | W U_{12} U_{23} U_{31}) \quad (13)$$

$$= I(V_1; Y_1 | W U_{12} U_{23} U_{31}) - I(V_1; V_2 V_3 X_2 X_3 | W U_{12} U_{23} U_{31}) \quad (14)$$

$$\leq I(V_1; Y_1 | W U_{12} U_{23} U_{31}) - I(V_1; X_2, X_3 | W U_{12} U_{23} U_{31}) \quad (15)$$

$$\leq I(V_1; Y_1 | W U_{12} U_{23} U_{31}) - I(V_1; X_2 + X_3 | W U_{12} U_{23} U_{31}) \quad (16)$$

$$= \sum_{\tilde{w}} P_{\tilde{W}}(\tilde{w}) \left[I(V_1; Y_1 | \tilde{W} = \tilde{w}) - I(V_1; S_1 | \tilde{W} = \tilde{w}) \right], \quad (17)$$

where the second equality follows from the second and the third Markov chain of Lemma 1. An astute reader can make the connection between the right hand side of the last equation and the capacity of the Gelfand-Pinsker channel [41]. Observe that the random variables appearing on the right hand side of the last equation, have the following probability mass function

$$P_{\tilde{W}} P_{V_1 | \tilde{W}} P_{X_1 | V_1, S_1, \tilde{W}} P_{Y_1 | X_1, S_1}$$

Consider the following binary Gelfand-Pinsker channel with $Y = X \oplus_2 S \oplus_2 N$, where X , S and N are binary valued, and $P(N = 1) = \delta$, $P(S = 1) = \alpha$. N and S are independent. Let $l : \{0, 1\} \rightarrow \mathbb{R}^+$ be a cost function with $l(0) = 0$ and $l(1) = 1$. Let $C(q, \alpha, \delta)$ denote the Gelfand-Pinsker capacity of this channel with the non-causal observation of the side information S at the encoder with cost constraint of q . Consider the following lemma.

Lemma 2. *For every triple $(q, \alpha, \delta) \in (0, \frac{1}{2})$, we have*

$$C(q, \alpha, \delta) < h_b(\delta * q) - h_b(\delta),$$

and equality holds if and only if any one of q, δ belongs to the set $\{0, \frac{1}{2}\}$ or $\alpha = 0$.

Using this lemma, we can see that for $q, \delta \in (0, \frac{1}{2})$, we have

$$R_1 < h_b(\delta * q) - h_b(\delta)$$

unless $H(X_2 + X_3 | \tilde{W}) = 0$. This along with the Markov chain $(X_2, V_2, V_{12}) - (W U_{23}) - (X_3, V_3, V_{13})$ of Lemma 1 implies that $H(X_2 | W U_{23} U_{12}) = H(X_3 | W U_{23} U_{31}) = 0$.

Now collecting all the results, we make the following closing arguments. If the first upper bound on R_1 , (see equation 9) as given by $I(V_1; Y_1|WU_{12}U_{23}U_{31}) - I(V_1; V_2V_3|WU_{12}U_{23}U_{31})$ is strictly smaller than $h_b(\delta * q) - h_b(\delta)$, then we are done, and $R_1 < h_b(\delta * q) - h_b(\delta)$.

If not, we have equality all the way from equation 13 to equation 17, and we must have

$$\begin{aligned} I(V_1; Y_1|\tilde{W}) &= I(V_1; Y_1|\tilde{W}, X_2, X_3) \\ &= I(X_1; Y_1|\tilde{W}, X_2, X_3) = H(X_1 \oplus_2 N_1|\tilde{W}, X_2, X_3) - H(N_1|\tilde{W}, X_2, X_3) \\ &= h_b(q * \delta) - h_b(\delta) \end{aligned}$$

Now looking at the fourth bound on R_1 (see equation 12), and using the fact that (WU_{23}) is independent of Y_2 and independent of Y_3 (see Lemma 1) we get

$$\begin{aligned} I(U_{12}U_{31}; Y_1|WU_{23}) - I(U_{12}; Y_2|WU_{23}) - I(U_{31}; Y_3|WU_{23}) &= I(X_2, X_3; Y_1|\tilde{W}) - I(X_2; Y_2) - I(X_3; Y_3) \\ &\leq 1 - H(X_1 \oplus_2 N_1|\tilde{W}, X_2, X_3) - 2 - 2h_b(\epsilon) \\ &= 2h_b(\epsilon) - h_b(\delta * q) - 1 \\ &< 0 \quad \text{if } h_b(\epsilon) < \frac{1 + h_b(\delta * q)}{2} \end{aligned}$$

Hence we have shown that if ϵ is such that

$$h_b(\delta * q) \leq h_b(\epsilon) < \frac{1 + h_b(\delta * q)}{2},$$

then $R_1 < h_b(q * \delta) - h_b(\delta)$. □

Proof. (Lemma 1): Substituting the relation $R_2 = K_2 + L_{12} + K_{23} + T_2$ in the 10th packing constraint we get

$$\begin{aligned} R_2 + K_1 + K_3 + K_{12} + L_{23} + S_2 &\leq I(WU_{23}U_{12}V_2; Y_2) + I(U_{12}; U_{23}|W) - S_{12} - S_{23} \\ &\leq I(WU_{23}U_{12}V_2; Y_2) \leq I(WU_{23}U_{12}U_{13}V_1V_2V_3X_2X_3; Y_2) \\ &= I(X_2; Y_2) \leq 1 - h_b(\epsilon) \end{aligned}$$

Since $R_2 = 1 - h_b(\epsilon)$, we must have equality everywhere in the above equation. Hence, we get $K_1 = K_3 = K_{12} = L_{23} = S_2 = 0$ and $S_{12} + S_{23} = I(U_{12}; U_{23}|W)$. Moreover, we have $(V_1, V_3, X_3, U_{13}) - (WU_{23}U_{12}V_2) - Y_2$. Since X_2 and Y_2 are related by a binary symmetric channel, using elementary probability argument, it can be easily shown that $(V_1, V_3, X_3, U_{13}) - (WU_{23}U_{12}V_2) - X_2$. Using a similar argument for the third receiver we get $K_1 = K_2 = K_{23} = L_{31} = S_3 = 0$, and $S_{23} + S_{31} = I(U_{23}; U_{31}|W)$, and $(V_1, V_2, X_2, U_{12}) - (WU_{23}U_{13}V_3) - (X_3, Y_3)$. Using these in the fourth covering constraint we get

$$\begin{aligned} I(U_{12}; U_{23}|W) + I(U_{23}; U_{31}|W) - S_{23} &= S_{12} + S_{23} + S_{31} \\ &\geq I(U_{12}; U_{23}|W) + I(U_{12}U_{23}; U_{31}|W) \end{aligned}$$

Hence we get

$$0 \leq S_{23} \leq -I(U_{12}; U_{31}|WU_{23}) \leq 0, \tag{18}$$

which implies that $S_{23} = 0$, and $U_{12} - (W, U_{23}) - U_{31}$. Substituting the condition that $S_2 = S_3 = 0$ in the 9th covering constraint gives us

$$\begin{aligned} 0 &= I(V_2; U_{31}|WU_{12}U_{23}) + I(V_3; U_{12}|WU_{13}U_{23}) + I(V_2; V_3|WU_{12}U_{23}U_{31}) \\ &= H(V_2|WU_{23}U_{12}) + H(V_3|WU_{23}U_{31}) - H(V_2, V_3|WU_{23}U_{12}U_{31}). \end{aligned}$$

This implies that $V_2 - (WU_{12}U_{23}) - (U_{31}V_3)$, and $V_3 - (WU_{31}U_{23}) - (U_{12}V_2)$. This relation along with $U_{12} - (W, U_{23}) - U_{31}$ gives us $(V_2, U_{12}) - (WU_{23}) - (V_3, U_{31})$. Using the above relations in the 12th and 15th packing constraints, we get

$$I(X_3; Y_3) = R_3 \leq I(U_{31}V_3; Y_3|WU_{23}) \leq I(X_3; Y_3)$$

and

$$I(X_3; Y_3) = R_3 \leq I(WU_{23}U_{31}V_3; Y_3|WU_{23}) \leq I(X_3; Y_3).$$

Combining these two equations, we get the constraint that (WU_{23}) is independent of Y_3 , and similarly independent of Y_3 . □

Proof. (Lemma 2): The Gelfand-Pinsker capacity $C(q, \alpha, \delta)$ is given by [27, 41]

$$C(q, \alpha, \delta) = \max_{P_{U|X|S}: E[X] \leq q} I(U; Y) - I(U; S)$$

where the maximization is over all conditional PMFs $P_{U|X|S}$ defined over $\mathcal{U} \times \{0, 1\}^2$ such that $E[X] \leq q$, where \mathcal{U} is a finite set and the joint distribution of the quadruple $(UXSY)$ is given by $P_S P_{U|X|S} P_{Y|XS}$. It is sufficient to restrict our attention to auxilliary alphabet \mathcal{U} of size 3. It is $|\mathcal{Y}| + |\mathcal{S}| - 1$.

The capacity of the above channel when both the encoder and the decoder has access to the side information is given by

$$C_B(q, \alpha, \delta) = \max_{P_{X|S}: E[X] \leq q} I(X; Y|S) = h_b(\delta * q) - h_b(\delta),$$

and there is a unique capacity achieving input distribution which is given by $P_{X|S}^*(0|0) = P_{X|S}^*(0|1) = 1 - q$. Hence the Gelfand-Pinsker capacity $C(q, \alpha, \delta)$ equals $C_B(q, \alpha, \delta)$ if and only if [42] the unique capacity achieving input distribution in the latter case can be expanded into a distribution Q_{UXSY} on the set $\mathcal{U} \times \{0, 1\}^3$ such that the following conditions are satisfied: (a) $|\mathcal{U}| = 3$, (b) X is a function of (U, S) , (c) $S - Y - U$, and (d) $U - (X, S) - Y$, and the marginal $Q_{XSY} = P_{XSY}^*$. We will show that there exists no such expansion by contradiction.

Let $\theta \triangleq q * \delta$, and without loss of generality let $\mathcal{U} = \{0, 1, 2\}$. Let there exist Q_{UXSY} for a triple $(q, \alpha, \delta) \in (0, \frac{1}{2})$ such that the four conditions are satisfied. Since $Q_{XSY} = P_{XSY}^*$, we have $Q_{SY}(0, 0) = (1 - \alpha)(1 - \theta)$, $Q_{SY}(0, 1) = (1 - \alpha)\theta$, $Q_{SY}(1, 0) = \alpha\theta$, and $Q_{SY}(1, 1) = \alpha(1 - \theta)$. Let $Q_{U|Y}(i|0) = \beta_i$, and $Q_{U|Y}(i|1) = \gamma_i$ for $i = 0, 1, 2$. From Q_{SY} and $Q_{U|Y}$ and imposing the Markov chain $S - Y - U$, we get the distribution of Q_{SYU} . Now since X is a function of (U, S) , let $Q_{X|US}(0|00) = z_0$, $Q_{X|US}(0|10) = z_1$, $Q_{X|US}(0|20) = z_2$, $Q_{X|US}(0|01) = z_3$, $Q_{X|US}(0|11) = z_4$, and $Q_{X|US}(0|21) = z_5$, where $z_i \in \{0, 1\}$ for $i = 0, 1, \dots, 5$. Using these we get the values for Q_{SYX} as given in Table 1.

S	Y	X	Q_{SYX}	P_{SYX}^*
0	0	0	$(1-\alpha)(1-\theta)[\beta_0 z_0 + \beta_1 z_1 + \beta_2 z_2]$	$(1-q)(1-\alpha)(1-\delta)$
0	0	1	$(1-\alpha)(1-\theta)[1 - \beta_0 z_0 - \beta_1 z_1 - \beta_2 z_2]$	$q(1-\alpha)\delta$
0	1	0	$(1-\alpha)\theta[\gamma_0 z_0 + \gamma_1 z_1 + \gamma_2 z_2]$	$(1-q)(1-\alpha)\delta$
0	1	1	$(1-\alpha)\theta[1 - \gamma_0 z_0 - \gamma_1 z_1 - \gamma_2 z_2]$	$q(1-\alpha)(1-\delta)$
1	0	0	$\alpha\theta[\beta_0 z_3 + \beta_1 z_4 + \beta_2 z_5]$	$(1-q)\alpha\delta$
1	0	1	$\alpha\theta[1 - \beta_0 z_3 - \beta_1 z_4 - \beta_2 z_5]$	$q\alpha(1-\delta)$
1	1	0	$\alpha(1-\theta)[\gamma_0 z_3 + \gamma_1 z_4 + \gamma_2 z_5]$	$(1-q)\alpha(1-\delta)$
1	1	1	$\alpha(1-\theta)[1 - \gamma_0 z_3 - \gamma_1 z_4 - \gamma_2 z_5]$	$q\alpha\delta$

Table 1: Distributions Q_{SYX} and P_{SYX}^* of (SYX)

Define

$$\psi_1 = \frac{1-q-\delta+q\delta}{1-q-\delta+2q\delta}, \quad \text{and} \quad \psi_2 = \frac{\delta-q\delta}{q+\delta-2q\delta}.$$

Equating $Q_{SYX} = P_{SYX}^*$ we get the following two equations.

$$\psi_1 = \beta_0 z_0 + \beta_1 z_1 + \beta_2 z_2 = \gamma_0 z_3 + \gamma_1 z_4 + \gamma_2 z_5$$

and

$$\psi_2 = \beta_0 z_3 + \beta_1 z_4 + \beta_2 z_5 = \gamma_0 z_0 + \gamma_1 z_1 + \gamma_2 z_2$$

We will see whether we can find z_i for $i = 0, 1, \dots, 5$ that satisfy the above two equations. First let us make two simple observations. The condition $\psi_1 = \psi_2$ is equivalent to the condition: $q = 1$ or $\delta = 0$ or $q = 1$. And the condition $\psi_1 + \psi_2 \leq 1$ is equivalent to the condition $q \geq \frac{1}{2}$, and equality holds in the latter if and only if either $q = 1$ or $\delta = 0$ or $\delta = 1$. Next we can see that there are four cases to consider.

Case 1: Two of $\{z_0, z_1, z_2\}$ and two of $\{z_3, z_4, z_5\}$ are zeros: We cannot have non-zero z_i s sharing the same β_i 's because of the first observation made above. So without loss of generality, let $z_1 = z_2 = z_3 = z_5 = 0$. Then $\psi_1 = \beta_0$ and $\psi_2 = \beta_1$. But, since $\beta_0 + \beta_1 \leq 1$, we get the condition that $q \geq \frac{1}{2}$, which is a contradiction.

Case 2: Two of $\{z_0, z_1, z_2\}$ and one of $\{z_3, z_4, z_5\}$ are zeros (or vice versa): Using the second observation made above, without loss of generality, let $z_1 = z_2 = z_5 = 0$. This implies that $\psi_1 = \beta_0$ and $\psi_2 = \beta_0 + \beta_1 = 1 - \beta_2$. Then $\beta_1 = 1 - \beta_2 - \beta_0 = \psi_2 - \psi_1$. Similarly, $\psi_2 = \gamma_0$ and $\psi_1 = \gamma_0 + \gamma_1$ which imply that $\gamma_1 = \psi_1 - \psi_2$. This implies that one of β_1 and γ_1 must be negative, unless $\psi_1 = \psi_2$ which leads to the condition $q = 0$ or $\delta = 0$ or $\delta = 1$. Hence a contradiction.

Case 3: One of $\{z_0, z_1, z_2\}$ and one of $\{z_3, z_4, z_5\}$ are zeros: Using the first observation made above, without loss of generality, let $z_2 = z_4 = 0$. Then $\psi_1 = \beta_0 + \beta_1 = 1 - \beta_2$ and $\psi_2 = \beta_0 + \beta_2 = 1 - \beta_1$, and similarly, $\psi_1 = 1 - \gamma_1$ and $\psi_2 = 1 - \gamma_2$. Hence we have $\gamma_1 = \beta_2$ and $\gamma_2 = \beta_1$. What do we do next? At this point, the Markov chain $U - (XS) - Y$ comes to our rescue. Let us look at the joint probability of (Y, U) conditioned on the event $(X, S) = (0, 0)$ as given in Table 2. Enforcing the Markov chain, we get $\frac{\beta_0}{\beta_1} = \frac{\gamma_0}{\gamma_1}$. This along with the relation $\beta_1 = \gamma_2$ and $\beta_2 = \gamma_1$, imply $\beta_1 = \beta_2$ or $\beta_1 + \beta_2 = 1$. The first gives us $\psi_1 = \psi_2$ and leads to $\delta = 0$ or $q = 1$ or $\delta = 1$. The second leads to $\delta = 0$ or $\delta = 1$ or $q = \frac{1}{2}$. Hence a contradiction.

Y	U	$(1 - q)Q_{YU S=0, X=0}$
0	0	$(1 - \theta)\beta_0$
0	1	$(1 - \theta)\beta_1$
1	0	$\theta\gamma_0$
1	1	$\theta\gamma_1$
0	2	0
1	2	0

Table 2: Conditional distribution of (Y, U) given $(S, X) = (0, 0)$.

Case 4: All of $\{z_0, \dots, z_5\}$ are zeros: In this case we get $\psi_1 = \psi_2 = 0$. This implies that $q = 1$ or $\delta = 1$ and $\delta = 0$ or $q = 1$. Hence a contradiction again.

□

References

- [1] K. Marton, “A coding theorem for the discrete memoryless broadcast channel,” *IEEE Trans. Inform. Theory*, vol. IT-25, no. 3, pp. 306–311, May 1979.
- [2] T. M. Cover, “Broadcast channels,” *IEEE Trans. Inform. Theory*, vol. IT-18, no. 1, pp. 2–14, Jan. 1972.
- [3] P. Bergmans, “Random coding theorem for broadcast channels with degraded components,” *Information Theory, IEEE Transactions on*, vol. 19, no. 2, pp. 197 – 207, mar 1973.
- [4] R. G. Gallager, “Capacity and coding for degraded broadcast channels,” *Probl. Peredachi Inf.*, vol. 10, no. 3, pp. 3 – 14, 1974.
- [5] P. Bergmans, “A simple converse for broadcast channels with additive white gaussian noise (corresp.),” *Information Theory, IEEE Transactions on*, vol. 20, no. 2, pp. 279 – 280, mar 1974.
- [6] E. C. Van der Meulen, “Random coding theorems for the general discrete memoryless broadcast channel,” *IEEE Trans. Inform. Theory*, vol. IT-21, no. 2, pp. 180–190, Mar. 1975.
- [7] B. E. Hajek and M. B. Pursley, “Evaluation of an achievable rate region for the broadcast channel,” *IEEE Trans. Inform. Theory*, vol. IT-25, no. 1, pp. 36–46, Jan. 1979.
- [8] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, New York, 1981.
- [9] S. I. Gelfand, “Capacity of one broadcast channel,” *Probl. Pered. Inform.*, vol. 13, no. 3, pp. 106108, JulySept. 1977; translated in *Probl. Inform. Transm.*, pp. 240242, JulySept. 1977.
- [10] T. M. Cover, “An achievable rate region for the broadcast channel,” *IEEE Trans. Inform. Theory*, vol. IT-21, no. 4, pp. 399–404, Jul. 1975.

- [11] P. P. Bergmans, "Random coding theorems for the broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 197–207, Mar. 1973.
- [12] —, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 279–280, Mar. 1974.
- [13] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 629–637, Nov. 1975.
- [14] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 60–64, Jan. 1977.
- [15] A. El Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 166–169, Mar. 1979.
- [16] K. Marton, "The capacity region of deterministic broadcast channels," in *Trans. Int. Symp. Inform. Theory*, Paris-Cachan, France, 1977.
- [17] M. S. Pinsker, "Capacity of noiseless broadcast channels," *Probl. Pered. Inform.*, vol. 14, no. 2, pp. 28–34, Apr.-Jun. 1978, translated in *Probl. Inform. Transm.*, pp. 97–102, Apr.-June 1978.
- [18] S. I. Gelfand and M. S. Pinsker, "Capacity of a broadcast channel with one deterministic component," *Probl. Pered. Inform.*, vol. 16, no. 1, pp. 24–34, Jan.-Mar. 1980, ; translated in *Probl. Inform. Transm.*, vol. 16, no. 1, pp. 17–25, Jan.-Mar. 1980.
- [19] A. El Gamal, "The capacity of the product and sum of two reversely degraded broadcast channels," *Probl. Pered. Inform.*, vol. 16, pp. 3–23, Jan.-Mar. 1980.
- [20] A. El Gamal and E. Van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-27, no. 1, pp. 120–122, Jan. 1981.
- [21] Y. Geng, C. Nair, S. Shamai, and Z. Wang, "On broadcast channels with binary inputs and symmetric outputs," *Proceedings of the International Symposium on Information Theory*, 2010. .
- [22] C. Nair and A. El Gamal "The Capacity Region of a Class of 3-Receiver Broadcast Channels with Degraded Message Sets," *IEEE Transactions on Information Theory*, vol. IT-55, no. 10, pp. 4479–4493, October, 2009.
- [23] A. A. Gohari and V. Anantharam, "Evaluation of marton's inner bound for the general broadcast channel," *IEEE Trans. Inform. Theory*, vol. 58, no. 2, pp. 608 – 619, Feb. 2012.
- [24] H. Sato, "An outer bound on the capacity region of broadcast channel," *IEEE Trans. on Inform. Theory*, vol. 24, pp. 374–377, May 1978.
- [25] C. Nair and A. E. Gamal, "An outer bound to the capacity region of the broadcast channel," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, July 2006.
- [26] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," *IEEE Transactions on Inform. Theory*, vol. 52, pp. 3936–3964, September 2006.

- [27] M. Costa, "Writing on dirty paper," *IEEE Trans. on Information Theory*, vol. 29, pp. 439–441, May 1983.
- [28] D. Krithivasan and S. S. Pradhan, "Distributed source coding using abelian group codes: A new achievable rate distortion region," *IEEE Trans. Inform. Theory*, March, 2011. .
- [29] A. Padakandla and S. S. Pradhan, "Nested linear codes achieve marton's inner bound for general broadcast channels", *Proc. IEEE Internatioanl Symposium on Information Theory (ISIT)*, July 2011.
- [30] A. Sahebi and S. Pradhan, "On the capacity of abelian group codes over discrete memoryless channels," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, 31 2011-aug. 5 2011, pp. 1743–1747.
- [31] —, "Codes over non-Abelian groups: Point-to-point communications and computation over MAC," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, Jul. 2012.
- [32] J. Korner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 219–221, March 1979.
- [33] T. S. Han and K. Kobayashi, "A dichotomy of functions $F(X,Y)$ of correlated sources (X,Y) ," *IEEE Trans. on Inform. Theory*, vol. 33, pp. 69–76, January 1987.
- [34] R. Ahlswede and T. S. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory," *IEEE Trans. on Inform. Theory*, vol. 29, pp. 396–412, May 1983.
- [35] T. Philosof, A. Kishty, U. Erez and R. Zamir, "Lattice Strategies for the Dirty Multiple Access Channel", *Proceedings of IEEE International Symposium on Information Theory*, July 2007, Nice, France.
- [36] G. Bresler, A. Parekh and D. Tse, "The Approximate Capacity of the Many-to-One and One-to-Many Gaussian Interference Channels" *IEEE Trans. Inform. Theory*, September 2012.
- [37] S. Shridharan, A. Jafarian, S. Vishwanath, and S. A. Jafar, "Lattice Coding for K User Gaussian Interference Channels", *IEEE Transactions on Information Theory*, April 2010. .
- [38] T. Berger, *Multiterminal Source Coding. In: The Information Theory Approach to Communications* (ed. G. Longo), *CISM Courses and Lecture Notes No. 229*. Springer, Wien-New York, 1977.
- [39] A. Padakandla and S. S. Pradhan, "A new achievable rate region for the multiple access channel with distributed states," In preparation.
- [40] C. Nair, "An achievable rate region for the 2-receiver broadcast channel obtained by viewing it as an interference channel," in *Proc. International Conference on Wireless Communications and Signal Processing*, Nanjing, China, 2009.
- [41] S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, pp. 19–31, 1980.

- [42] S. S. Pradhan, J. Chou, and K. Ramchandran, “Duality between source coding and channel coding and its extension to the side information case,” *IEEE Trans. Inform. Theory*, vol. 49, pp. 1181–1203, May 2003.